Federated Learning for Generalizable Litter Detection in Smart Cities

Luciano Baresi*, Simone Bianco+, Livia Lestingi*, and Iyad Wehbe*
*Politecnico di Milano, DEIB, Milan, Italy, {name.surname}@polimi.it
+Università di Milano Bicocca, DISCO, Milan, Italy, simone.bianco@unimib.it

***Abstract*—Abandoned garbage in unauthorized urban areas significantly affects the environmental sustainability and the quality of life of the community. Automated garbage detection methods are essential to initiate garbage disposal processes, but require accurate image processing models. However, traditional ML-based techniques raise privacy concerns. In this context, Cobol envisages the collaboration of multiple municipalities and takes advantage of a Federated Learning (FL) setting to train accurate litter detection models without sharing sensitive citizen data. This extended abstract presents the assessment of a FL framework to train the YOLOv8-nano object detection model for images tested on realistic heterogeneous splits of TACO and PlastOPol, which are two benchmark datasets of garbage images.**

I. Introduction

The illegal disposal of garbage is a critical challenge for modern urban communities as it negatively affects public health and the quality of life of citizens. Detecting garbage accurately is essential for waste management, as it can be used to inform authorities on which locations require immediate intervention. Although recent advances in computer vision have significantly improved automated litter detection systems [1], these are typically based on extensive aggregated data sets. However, centralized data collection processes raise concerns about privacy, data ownership, and citizen trust.

Federated Learning (FL) is a decentralized machine learning paradigm where models are trained by multiple clients [2]. In FL, training occurs in rounds. During each round, a global server distributes the current model parameters to multiple client nodes. Clients train the model locally on their own private datasets for several epochs and subsequently send updated model parameters (not the raw data) back to the server.

COmmunity-Based Organized Littering (COBOL) aims to develop a community-driven data-centric system (represented in Fig. 1) to detect and address garbage in urban and suburban areas [3]. Citizens contribute via geo-located photo reports or passive on-device litter detection. In both cases, the system analyzes images to confirm the presence of waste and classifies its type and size. The system then informs local authorities and triggers appropriate cleanup actions.

This work investigates the potential of FL for generalizing across different datasets in the detection of trash from photos taken by citizens. The evaluation exploits two data sets with significantly different visual and contextual properties. One is TACO[1], which contains various environmental scenes such as forests, beaches, and urban streets. The second is PlastOPol [4], which contains images collected by the Marine Debris Tracker with a total of 5300 litter instances.

This extended abstract introduces our methodological framework, highlights key experimental results demonstrating the effectiveness of FL-based solutions, and discusses implications for practical deployment within smart cities. Our findings indicate that FL trades accuracy —compared to traditional centralized training solutions— for the privacy of participants.

II. Experiment Setup

We propose a FL solution that specifically targets garbage detection. We selected YOLOv8-nano [5], an object detector that makes predictions based on a bounding box to identify the position of the litter in the image. This lightweight yet powerful model is suitable for deployment on mobile devices. The implemented FL framework exploits Flower[2], a well-known framework for the effective simulation capabilities of FL-based systems and a user-friendly interface.

The parameters shared by the clients are typically aggregated using a refined version of the Federated Averaging (FedAvg) algorithm [2]. Our solution prioritizes updates from clients with larger datasets and greater local improvements, facilitating model stability under heterogeneous conditions.

For training and testing datasets, we used two benchmark data sets, TACO and PlastOPol, both unified under a 'litter' format of one class for consistency. Data were distributed across 2, 4, 8, and 16 clients to simulate realistic FL deployments. In the 2-client setup, each client received one full dataset. We applied two partitioning strategies for larger federations: random splitting (yielding homogeneous data distributions) and clustering based on visual similarity using InceptionV3 features and k-means (yielding heterogeneous, imbalanced splits). The latter strategy represents a realistic setting in which higher visual similarity is expected among images taken in the same municipality (thus, stored in the same node of the federation).

To provide a baseline for comparison, we also trained three centralized models: one using the combined TACO and PlastOPol datasets, and two using each dataset individually. These centralized models were trained for the same number of total epochs as the federated scenarios. Their test accuracy on each centralized experiment was used as a reference to assess the performance and generalization of the federated models.

[1] Available at: http://tacodataset.org.

[2] Available at: https://flower.ai.

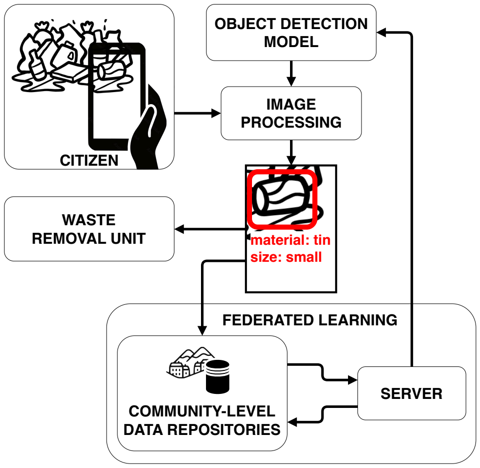

Fig. 1: Overview of the COBOL approach.

III. Results

We evaluate the resulting model accuracy using mAP@50 (mean Average Precision at the intersection over union between predicted and groundtruth bounding boxes equal to 0.5) [6]. We trained YOLOv8n in seven FL configurations involving 2, 4, 8, and 16 clients, using both random (r) and visual similarity-based heterogeneous (h) data splits. For evaluation, we used three separate test sets: one from TACO, one from PlastOPol, and a third that combines samples from both. Each experiment was repeated five times to ensure robust evaluation and avoid results influenced by randomness.

The results are reported in Table I and show the performance of each configuration in terms of average mAP@50 over five independent repetitions. For comparison, we also report the baseline performance obtained by centralized training on just one dataset and the upper bound performance obtained by centralized training on both datasets. The 2-client FL configuration, where each client trains on one complete dataset, achieves a mAP@50 score that exceeds the centralized baselines when tested cross-dataset (the model is tested on a dataset that was not seen during local training by any client). However, as the number of clients increases, performance drops due to data fragmentation, particularly under heterogeneous splits. This highlights the need to incorporate data heterogeneity management techniques, such as clustering clients based on data distribution. FL configurations based on visual similarity clustering generally outperform random ones, confirming the importance of data distribution strategies to maintain model accuracy in federated settings.

IV. Conclusions

We propose a FL approach to train a garbage detection model as part of the solution envisaged in the context of COBOL. The application assumes that data are distributed across municipalities, each with different environmental conditions and types of litter, leading to heterogeneous datasets. FL enables collaborative model training while keeping citizen data local and preserving privacy and data ownership. Experiments show that FL models achieve accuracy levels comparable to centralized training, especially when client data are structured and minimally fragmented.

We aim to explore a two-level federated architecture, enabling the deployment of the presented FL solution in a small-scale pilot. At the first level, models are trained on citizen devices and aggregated on municipal servers. At the second level, municipality-level models are aggregated by a central server, enabling broader knowledge sharing while maintaining privacy and locality at each layer. We also plan to assess the transferability of our findings to different object detection tasks within the smart city domain (e.g., urban decay detection).

TABLE I: Average mAP@50 scores per configuration and test dataset.

Nodes	Data Split	All	TACO	Plasto
1 (upper bound)	-	0.679	0.543	0.786
1 (baseline)	only TACO	0.566	0.511	0.616
1 (baseline)	only PlastoOpol	0.625	0.454	0.759
2	-	0.619	0.468	0.741
4	h	0.473	0.331	0.588
4	r	0.587	0.449	0.699
8	h	0.587	0.436	0.688
8	r	0.550	0.376	0.677
16	h	0.523	0.398	0.628
16	r	0.451	0.336	0.547

Acknowledgements

This work is supported by COBOL, funded by the MUR under the PRIN 2022 PNRR program (nr. P20224K9EK).

References

[1] I. Dawar, A. Srivastava, M. Singal, N. Dhyani, and S. Rastogi, "A systematic literature review on municipal solid waste management using machine learning and deep learning," *Artificial Intelligence Review*, vol. 58, no. 6, p. 183, 2025.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[3] L. Baresi, S. Bianco, A. Di Salle, L. Iovino, L. Mariani, D. Micucci, L. B. R. dos Santos, M. T. Rossi, and R. Schettini, "COBOL: COmmunity-based Organized Littering," in *Euromicro Conference on Software Engineering and Advanced Applications*. IEEE, 2024, pp. 511–517.

[4] M. Córdova, A. Pinto, C. C. Hellevik, S. A.-A. Alaliyat, I. A. Hameed, H. Pedrini, and R. d. S. Torres, "Litter detection with deep learning: A comparative study," *Sensors*, vol. 22, no. 2, p. 548, 2022.

[5] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[6] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.